

# Social media data as a proxy for hourly fine-scale electric power consumption estimation

Environment and Planning A:

Economy and Space

0(0) 1–5

© The Author(s) 2018

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0308518X18786250

journals.sagepub.com/home/epn



**Chengbin Deng**  and **Weiying Lin**

State University of New York at Binghamton, USA

**Xinyue Ye**

Kent State University, USA

**Zhenlong Li**

University of South Carolina, USA

**Ziang Zhang and Ganggang Xu**

State University of New York at Binghamton, USA

## Abstract

Accurate forecasting of electric demand is essential for the operation of modern power system. Inaccurate load forecasting will considerably affect the power grid efficiency. Forecasting the electric demand for a small area, such as a building, has long been a well-known challenge. In this research, we examined the association between geotagged tweets and hourly electric consumption at a fine scale. All available geotagged tweets and electric meter readings were retrieved and spatially aggregated to each building in the study area. Comparing to traditional studies, the usage of geotagged tweets is to reflect human activity dynamics to some degree by considering human beings as sensors, which therefore can be employed at the building level. High correlation is found between the human activity indicator and the power consumption as supported by a correlation coefficient level over 0.8. To the best of our knowledge, rare studies placed an emphasis on hourly electric power consumption using social media data, especially at such a fine scale. This research shows the great potential of using Twitter data as a proxy of human activities to model hourly electric power consumption at the building level. More studies are warranted in the future to further examine the effectiveness of the proposed method in this research.

---

## Corresponding author:

Xinyue Ye, Department of Geography, Kent State University, Kent, OH 44240, USA.

Email: xye5@kent.edu

**Keywords**

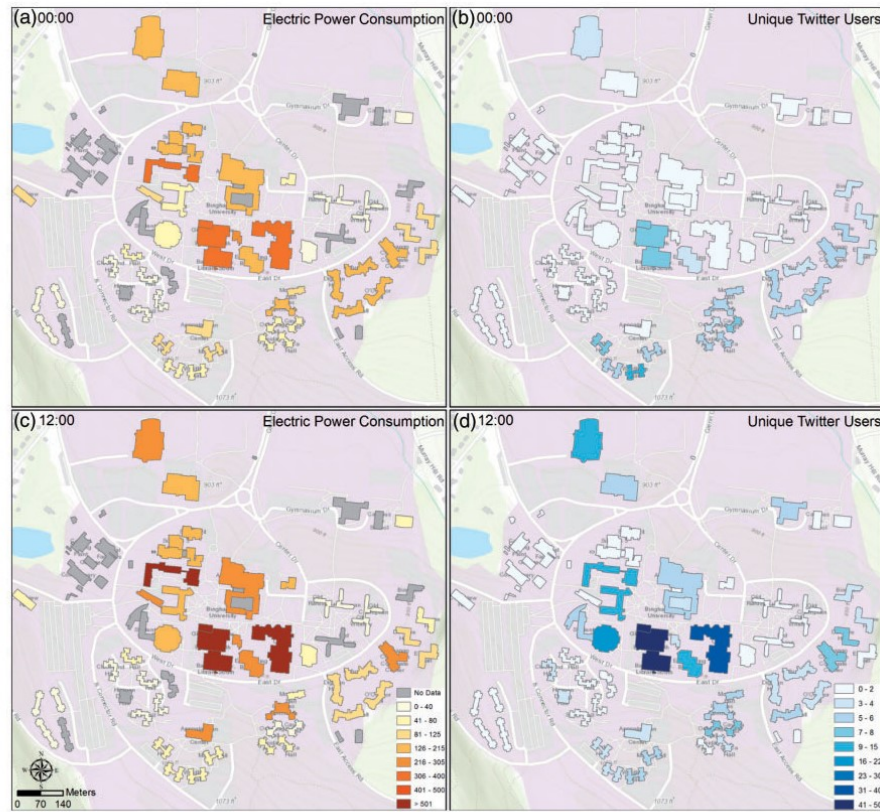
Social media, geotagged Twitter, electric power consumption, electric load forecasting, building-level prediction

The operation of modern power system relies on accurate forecasting of electric demand. Inaccurate load forecasting could have a negative impact on the power grid efficiency. Moreover, the uncertainty of renewable energy generation requires power engineers have a better understanding of the load profile. Similarly, hourly based building load forecasting information could enable the building operator to perform much efficient control and energy storage systems (Zhao and Magoulès, 2012). Forecasting the electric demand for a small area, such as a building, has been a well-known challenge (Hippert et al., 2001).

Conventional machine learning-based methods mainly rely on the temporal correlation of energy usage. Such methods work for large-scale load estimation but not for the building-level load forecasting, as less pattern could be detected in a smaller area. At the same time, scholars also employ nighttime light (NTL) satellite images to estimate electric consumption. However, the use of NTL image is still insufficient for load estimation. Spatially, popular NTL data with a coarse spatial resolution (e.g. 1-km DMSP OLS imagery, and 750-m VIIRS DNB imagery) are only suitable for electric consumption estimation at national, continental, or global level (Elvidge et al., 1997), and cannot provide the estimates at a finer scale. Temporally, an NTL image at one point in time is not applicable to estimate hourly electric consumption. New features beyond temporal information are needed to improve the performance of load forecasting.

To address these limitations, this study examined the performance of using new datasets with both higher spatial and temporal resolutions for modeling fine-scale electric consumption. Specifically, we employed geotagged tweets to estimate hourly electric consumption at the building level on the campus of State University of New York at Binghamton. All tweets posted between mid-January and mid-May (in accordance with 2015 Spring semester) were extracted by using Twitter API. The electric load consumption of each building on campus was retrieved from smart meters. Geotagged Twitter data were aggregated to each building and averaged to each hour of weekday. Weekend data were excluded when no class runs. Such processing steps were also applied to the electric consumption data. Figure 1 shows the spatial pattern of unique Twitter users and electric consumption of each building on campus at midnight and noon. It can be observed that academic buildings (larger ones at the center of each figure) have both higher amounts of electric consumption and unique Twitter users than student housing buildings (at the bottom and lower right of each figure), regardless of the hour of day. Besides, temporal change can be found with both variables in buildings on campus, but their patterns seem to be different in academic buildings and dormitory.

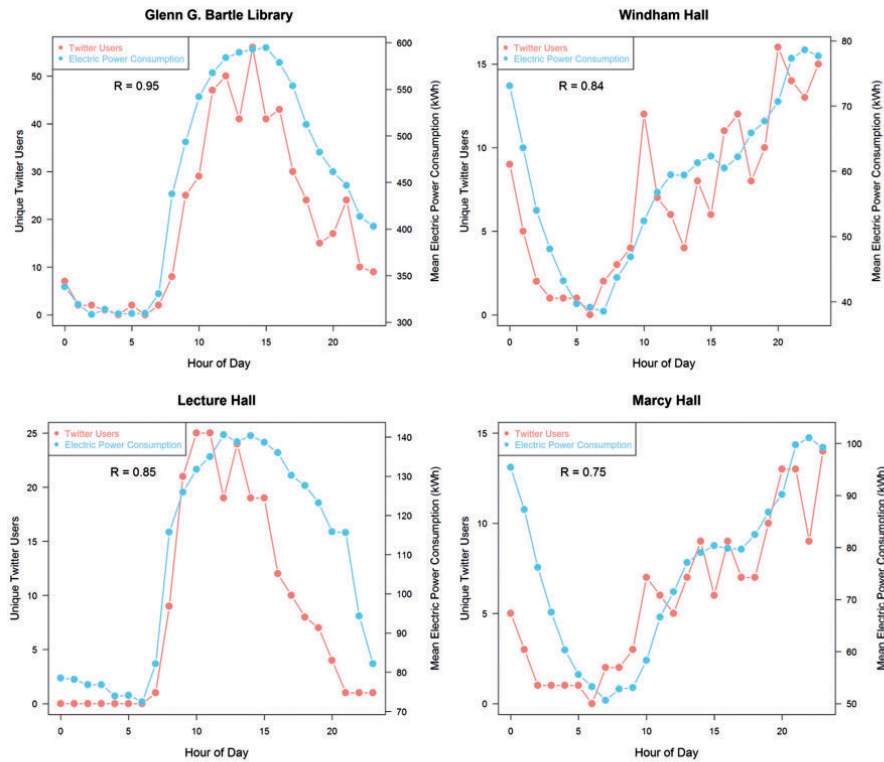
To further study their relationship, Figure 2 illustrates their comparison in four representative buildings at different time. Plots in the first column show the hourly pattern of two academic buildings. The electric consumption peak of academic buildings is approximately at noon, and the valley at approximately 4 am in the morning. This pattern also applies to the red curve of unique Twitter users, which is supported by a Pearson's correlation coefficient of approximately 0.9. The second column shows the hourly patterns of two student housing buildings. Comparatively, the peak and valley of electric consumption and unique Twitter users in dormitory occur at midnight and 6 am in the morning, respectively.



**Figure 1.** Spatial distribution of electric power consumption and unique Twitter users on campus at different time of day. Twitter, created in March 2006, is an online social networking platform where users post and interact with each other through messages with limited characters (called tweets).

The Pearson's correlation coefficient is approximately 0.8, which also suggests a high association between electric consumption and unique Twitter users in dormitory ( $P < 0.01$  for correlation analysis above). While Twitter itself hardly contributes to power usage, more tweets imply a higher chance of more people in a building, and consequently, more heating, air conditioning, lighting, as well as other human activities may require more power.

Geotagged Twitter data have been widely used in a variety of spatial studies, such as urban vitality and human mobility (Shaw et al., 2016). However, rare studies placed an emphasis on hourly fine-scale electric power consumption. The essence of geotagged tweets is that such dataset can reflect human activity dynamics to some degree by considering human beings as sensors (Goodchild, 2007). This research shows the great potential of using Twitter data as a proxy of human activities to model hourly electric power consumption at the building level. While we performed our analysis on campus, the load patterns of those academic and business buildings tend to be similar to those of office buildings in central business area. More studies are warranted in the future to further examine the effectiveness of the proposed method in this research.



**Figure 2.** Curves of the amount of unique tweet users (red) and average hourly electric power consumption (blue) in different buildings on campus. Left column shows curves of academic buildings, and right column shows those of dormitory.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research was partially funded by an Interdisciplinary Collaboration Grant at the State University of New York at Binghamton (Grant number: 1139728) and National Science Foundation (Grant number: 1637242 and 1739491).

### ORCID iD

Chengbin Deng  <http://orcid.org/0000-0002-5459-5586>

## Reference

- Elvidge CD, Baugh KE, Kihn EA, et al. (1997) Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing* 18(6): 1373–1379.
- Goodchild MF (2007) Citizens as sensors: The world of volunteered geography. *GeoJournal* 69(4): 211–221.
- Hippert HS, Pedreira CE and Souza RC (2001) Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems* 16(1): 44–55.
- Shaw SL, Tsou MH and Ye X (2016) Human dynamics in the mobile and big data era. *International Journal of Geographical Information Science* 30(9): 1687–1693.
- Zhao HX and Magoulès F (2012) A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*. Pergamon. Available at: <https://doi.org/10.1016/j.rser.2012.02.049> (accessed 19 June 2018).